Cancer Causes and Control, 1997, 8, pp. 544-553

A case-control study of cytochrome *P450 1A1*, glutathione S-transferase M1, cigarette smoking and lung cancer susceptibility (Massachusetts, United States)

Montserrat García-Closas, Karl T. Kelsey, John K. Wiencke, Xiping Xu, John C. Wain, and David C. Christiani

(Received 14 August 1996; accepted in revised form 22 January 1997)

Cytochrome *P450 1A1* (*CYP1A1*) and glutathione S-transferase M1 (*GSTM1*) genetic polymorphisms are involved in the activation and detoxification of chemical carcinogens found in tobacco smoke; thus they may influence host susceptibility to lung cancer. In this study at Massachusetts General Hospital (Boston, MA, USA) of 416 cases and 446 controls (mostly White) we evaluated the association between the *CYP1A1 MspI* and *GSTM1* polymorphisms and lung cancer risk, and their interaction with cigarette smoke. The *CYP1A1 MspI* heterozygous genotype was present in 18 percent of cases and 16 percent of controls, and one percent of cases and controls were *CYP1A1 MspI* homozygous variant. The *GSTM1* null genotype was detected in 54 percent of cases and 52 percent of controls. After adjusting for age, gender, pack-years of smoking, and years since quitting smoking, while neither the *CYP1A1 MspI* heterozygous genotype alone nor the *GSTM1* null genotype alone were associated with a significant increase in lung cancer risk, having both genetic traits was associated with a twofold increase in risk (95 percent confidence interval [CI] = 1.0-3.4). Our data did not provide enough evidence for a substantial modification of the effect of pack-years on lung cancer risk by the *CYP1A1 MspI* and *GSTM1* genotypes. However, limitations of our study preclude a conclusion about this potential interaction. *Cancer Causes and Control* 1997, 8, 544-553

Key words: Cytochrome P450, genotype, glutathione transferases, lung neoplasm, smoking, United States.

Introduction

Most carcinogens in tobacco smoke require metabolic activation by phase I enzymes in order to manifest their carcinogenic effects. The activated intermediates resulting from these reactions then can be detoxified through reactions catalyzed by phase II enzymes.[1,2] Therefore, genetically determined differences in metabolizing

Authors are with the Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA (Dr Garcia-Closas); Department of Environmental Health, Occupational Health Program, Harvard School of Public Health, Boston, MA (Drs Kelsey, Xu, Christiani); Department of Cancer Biology, Harvard School of Public Health, Boston, MA (Dr Kelsey); Laboratory for Molecular Epidemiology, Department of Epidemiology and Biostatistics, University of California at San Francisco, CA, USA (Dr Wiencke); Thoracic Surgery Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA (Dr Wain); Pulmonary and Critical Care Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA (Dr Christiani). Work was performed at the Departments of Epidemiology and Environmental Health, Harvard School of Public Health, Boston, MA. Address correspondence to Dr Christiani, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115, USA. *This study was supported by Grant ES/CA P01 06409 and ES00002 from the US National Institute of Environmental Health Sciences and the National Cancer Institute. Dr García-Closas was supported by a scholarship from the Real Colegio Complutense, Cambridge, MA.*

enzymes involved in these reactions could influence host susceptibility to lung cancer, as well as modify the relationship between cigarette smoking and lung cancer risk.[3]

Cytochrome *P450 1A1* and glutathione S-transferase M1 (*GSTM1*) are, respectively, a phase I and phase II enzyme involved in the metabolic activation and detoxification of polyaromatic hydrocarbons (PAH) and other combustion products found in cigarette smoke.[2,3] These two enzymes have been found to be polymorphic in human populations. There are two described restriction fragment-length polymorphisms (RFLP) of the gene coding for *P450 1A1*: the *MspI* [4] and the *Ile/Val* [5] polymorphisms. While the relationship between these two closely linked polymorphisms and the enzyme activity remains to be established,[4,6,7] the homozygous variant genotype of both polymorphisms has been associated with an increased risk of lung cancer in Japanese populations.[8-11] In contrast, this association has not been confirmed in studies conducted in Caucasian (White) populations.[12-15] The polymorphism in the *GSTM1* gene is due to a complete deletion of the gene which results in a lack of enzyme activity.[16-18] Several epidemiologic studies[19-21] have evaluated the association between lung cancer and *GSTM1* null genotype (homozygous deletion) obtaining contradictory results. Overall, these studies indicate that subjects with the *GSTM1* null genotype may have a 10 to 60 percent average increase in lung cancer risk. This increase in risk has been stronger and found more consistently in Japanese than in Caucasian populations.

Several studies have evaluated the potential modification of the association between these polymorphisms and lung cancer by cigarette smoke.[8,11,13,14,20,22-24] However, results from these studies have been contradictory and the question of whether these genes have stronger associations with lung cancer at high or low levels of exposure, or whether the associations are the same across levels of exposure remains controversial. The small number of subjects included in most of these studies, the use of different methods of assessment and categorization of cumulative dose, and differences in ethnicity are possible explanations for the apparently contradictory results.

Some studies also have evaluated the potential interaction between the *CYP1A1* and *GSTM1* genotypes. Japanese studies have reported that subjects with the combined *GSTM1* null genotype and *CYP1A1 MspI* or *Ile/Val* polymorphisms are at remarkably high risk of lung cancer, especially at low cumulative doses of cigarette smoke.[10] Two subsequent Scandinavian studies[25,26] suggested an interaction between the *GSTM1* null genotype and the *CYP1A1 MspI* variant allele or the aryl hydrocarbon hydroxylase (AHH) inducible phenotype, however these studies were too small to be conclusive.

We recently reported[27] a positive association between the *CYP1A1 MspI* allele and lung cancer risk, which was only evident after adjusting for cumulative smoking dose. Specifically, in a case-control study of 207 lung cancer cases and 283 controls from a North American (mostly White) population, the combined homozygous/heterozygous *MspI* variant of the *CYP1A1* gene was associated with a twofold increase in lung cancer risk (95 percent confidence interval [CI] = 1.2-3.7). We have now expanded our study to include 416 cases and 446 controls, providing us with one of the largest studies conducted to date. In this expanded population we have evaluated the association between the *CYP1A1 MspI* and *GSTM1* polymorphisms and lung cancer risk, as well as the modification of the effect of cigarette consumption by these polymorphisms.

Materials and methods

The present study is a case-control study of incident cases of lung cancer which was conducted at the Massachusetts General Hospital (MGH) in Boston, MA, United States. Eligible cases included all patients with newly diagnosed primary lung cancer (stages I and II) presenting for thoracic surgery at the MGH between December 1992 and August 1996, and who had a confirmed diagnosis after histologic examination of post-surgery tumor samples. Lung cancer patients were referred by physicians from various services of the MGH as well as from physicians outside MGH, and the vast majority (~ 90 percent) lived in the greater Boston area.

During the study period, we identified 572 eligible patients, most of whom were White. Among these, 42 refused to participate, 34 were missed, 26 were unable to participate, and 11 agreed to participate but did not complete the questionnaires. This left 459 lung cancer cases with a participation rate of 80 percent. Controls (n = 458) were selected among friends or spouses of lung cancer patients (n = 167), and among friends or spouses of cardiac surgery patients (n = 83) or other thoracic surgery patients (n = 208), all from MGH. A formal record of the participation rate among controls was not kept during the most of the study period. However, based on an informal record of potential candidates identified by the hospital nurse, our best estimate of the participation rate among controls is about 90 percent.

Data collection

Stage and histologic type of lung cancer was assessed by histologic examination of post-surgery tumor samples. Information on demographic characteristics (gender; age; level of education; race; and family history of cancer in parents, siblings, or children) was collected through questionnaires administered to cases and controls by trained interviewers at the hospital. Smoking and other occupational/environmental exposure history was collected through a modified standardized American Thoracic

Society respiratory questionnaire.[28] This questionnaire included information on: current smoking status; age when started smoking; whether they temporarily quit smoking for six months or more during the period they smoked; age when quit smoking; and the average amount of cigarettes smoked per day, period, and frequency of exposure to other environmental or occupational substances such as asbestos and solvents. Levels of food intake were assessed through a semiquantitative food-frequency questionnaire with 126 food items.[29] Details on the assessment of dietary intake in the population under study have been previously described.[30] Questions about current habits were referred to time of diagnosis for cases and time of interview for controls.

Blood samples were obtained in the hospital from cases and controls and sent to the molecular biology laboratory for genotype determination. DNA was obtained from heparinized whole blood by use of *Chelex*® solution as described by Walsh *et al.* [31] Genotyping for *CYP1A1 MspI* and *GSTM1* was completed using PCR-based methods previously published.[27,32,33]

Data analysis

Odds ratios (OR) and CIs were used to quantify the association between lung cancer risk and genotype. Generalized additive models were used to assess non-linear associations of continuous covariates on the logit of the probability of disease.[34] The following potential confounders were considered in the analysis: gender (male, female); race (White and others); level of education (less than high school, high school graduate, 1-4 years of college, and college graduate); smoking status (never, current or quit within the last year, former); current age; age at onset of smoking; years since quit smoking; average duration of smoking; average cigarettes per day; pack-years smoked; occupational exposure to asbestos (yes, no); family history of lung cancer among parents or siblings (yes, no); and dietary intake of antioxidant vitamins (vitamin C, vitamin E, alpha-carotene and beta-carotene). Among these potential confounders, only age, gender, dose and duration of cigarette smoking, and years since quit smoking were included in our final model (see Results section).

Generalized additive models indicated the presence of a non-linear association on the logit of the probability of disease for age and pack-years, and a linear association for years since quit smoking. The functional form for age and pack-years was approximated by second-degree polynomials. Therefore, the adjusted genotype ORs were obtained from parametric models which included age and pack-years as second-degree polynomials, a linear term for years since quit smoking, and an indicator term for gender. The genotype ORs were practically unchanged by alternative ways of modeling dose and duration of cigarette smoking, *e.g.*, including dose and duration as two separate terms, or including pack-years only. Statistical analysis were performed with the statistical software STATA version 5.0 and S-Plus version 3.3 for Windows.

Results

The study population included 459 cases of lung cancer and 458 controls. Five controls and 37 cases were excluded because we were unable to obtain a blood sample for genotype analysis. Among the remaining subjects, six cases and seven controls were excluded because of incomplete smoking data, leaving 416 cases and 446 controls available for the main analysis.

The distribution of characteristics of cases and controls with complete genotype and smoking information (n = 917) is displayed in Table 1. Approximately, 97 percent of cases and 98 percent of controls were Whites. Respectively, cases and controls had a percentage of females of 46 percent and 53 percent, and a median age of 67 years (33 to 89 range) and 64 (27 to 84 range) years. Of the cases, current smokers accounted for 41 percent, former smokers (defined by having quit smoking at least one year before enrollment) for 54 percent, and never-smokers for five percent. Of the controls, current smokers accounted for 17 percent, former smokers for 52 percent, and never-smokers for 31 percent. The most common histologic type was adenocarcinoma (54 percent), followed by squamous cell carcinoma (28 percent), large cell carcinoma (five percent), small cell carcinoma (four percent), and others including mixed cell types (eight percent).

Association between CYP1A1 MspI and GSTM1 genotypes and lung cancer

The distribution of the *CYP1A1 MspI* and *GSTM1* genotypes was very similar for cases and controls (Table 1). The *CYP1A1 MspI* homozygous variant genotype was present only in four controls and four cases. Since past studies have shown different magnitude of associations for homozygous variant and heterozygous subjects[9,10] these eight subjects were excluded from further analyses.

A positive association between the *CYP1A1 MspI* polymorphism and lung cancer risk has been reported previously in a subset of the population under study.[27] This association was only present after adjusting for cigarette smoking. In the current population, the *CYP1A1 MspI* heterozygous genotype was associated with a 50 percent increase in risk of lung cancer (CI = 1.0-2.3) (Table 2). As in our pervious report, pack-years was the main variable responsible for the differences between the crude and adjusted estimates. This reflected a negative association between the average cigarettes smoked per day and the *CYP1A1 MspI* heterozygous genotype among the control group (23 ± 15 and 17 ± 11 [mean ±

Table 1. Distribution of selected variables among lung cancer patients and controls, Massachusetts General Hospital

Characteristics	Cases (*n* = 416)	Controls (*n* = 446)
Age in years, median (range)	67 (33-89)	64 (27-84)
Gender, *n* (%) females	193 (46.4)	237 (53.1)
Race, *n* (%) White	403 (96.9)	436 (97.8)
Education[a] *n* (%) College graduate	86 (21.3)	96 (21.8)
Smoking status, *n* (%)		
Never-smokers	21 (5.0)	139 (31.2)
Former smokers	226 (54.3)	233 (52.2)
Current smokers	169 (40.6)	74 (16.6)
Duration of smoking in years, mean (SD)		
Former smokers	36 (12)	24 (12)
Current smokers	44 (11)	38 (11)
Average cigarettes per day, mean (SD)		
Former smokers	30 (16)	22 (15)
Current smokers	29 (15)	21 (11)
Years since quit smoking, mean (SD)	13 (10)	20 (12)
Passive exposure to cigarette smoke at home or work[b]		
n (%) positive	18 (94.7)	127 (92.7)
Occupational exposure to asbestos,[c] *n* (%) positive	60 (14.6)	43 (9.6)
CYP1A1 MspI		
n (%) homozygous wild-type	337 (81.0)	369 (82.7)
n (%) heterozygous	75 (18.0)	73 (16.3)
n (%) homozygous variant	4 (1.0)	4 (1.0)
GSTM1		
n (%) homozygous wild-type or heterozygous	190 (45.7)	214 (48.0)
n (%) homozygous deleted	226 (54.3)	232 (52.0)
Family history of lung cancer,[d] *n* (%) positive	81 (22.8)	58 (15.1)
Histologic type[e]		
Adenocarcinoma, *n* (%)	225 (54.3)	—
Squamous cell carcinoma, *n* (%)	116 (28.1)	—
Small cell carcinoma, *n* (%)	18 (4.3)	—
Large cell carcinoma, *n* (%)	23 (5.6)	—
Other tumors,[f] *n* (%)	32 (7.7)	—

[a] Information was missing for 13 cases and 5 controls.
[b] Among never-smokers only. Information was missing for 2 cases and 2 controls.
[c] Information was missing for 5 cases.
[d] Information was missing for 61 cases and 61 controls.
[e] Information was missing for 2 cases.
[f] Includes 14 cases with mixed tumors and 18 cases with more than one tumor.
SD = standard deviation.

standard deviation] cigarettes per day for m1/m1 and m1/m2 respectively; Wilcoxon Sum Rank Test: $P = 0.01$). In contrast to the *MspI* genotype, our data did not show evidence for an association between the *GSTM1* null genotype and lung cancer risk (Table 2). When different histologic types of lung cancer were evaluated separately, no relevant differences in risk were observed for either of the genotypes (data not shown).

Models restricted to Whites or including dummy variables for different levels of education did not show important changes in the estimated genotype ORs. We also evaluated the confounding effects of family history of lung cancer, dietary intake of antioxidant vitamins and occupational exposure to asbestos among the subset of subjects with complete information on these variables ($n = 684$). Since we did not find evidence for important confounding by either of these variables and given that several subjects had missing information, we did not include them in the final adjusted model.

Interaction between CYP1A1 MspI, GSTM1, and cigarette smoke

Table 3 shows results from a stratified analysis performed to evaluate the modification of the association between

Table 2. Association between lung cancer risk, *CYP1A1 MspI* and *GSTM1* polymorphisms

	Cases	Controls	Crude		Adjusted[b]	
			OR	(CI)[a]	OR	(CI)[a]
CYP1A1 MspI [c]						
m1/m1	337	369	1.0	—	1.0	—
m1/m2	75	73	1.1	(0.8-1.6)	1.5	(1.0-2.3)
GSTM1 [d]						
Present	190	214	1.0	—	1.0	—
Null	226	232	1.1	(0.8-1.4)	1.0	(0.7-1.4)

[a] OR = odds ratio; (CI) = 95 percent confidence interval.
[b] Odds ratios adjusted for age, gender, pack-years, years since quit smoking, and the other genotype.
[c] m1/m1 = homozygous wild-type; m1/m2 = heterozygous. Homozygous variant subjects (m2/m2) were excluded from this analysis ($n = 8$).
[d] *GSTM1* null genotype refers to homozygous deletion.

Table 3. Association between lifetime cumulative cigarette dose and lung cancer risk stratified by the *CYP1A1 MspI* and *GSTM1* genotypes[a]

Pack-years of smoking	Cases	Controls	Adjusted[b]	
			OR	(CI)[a]
All subjects				
≤ 20	46	119	1.0	—
21-40	76	102	1.2	(0.7-1.9)
41-60	109	50	3.0	(1.7-5.3)
> 60	161	36	5.8	(3.1-10.6)
CYP1A1 MspI genotype[c,d]				
m1/m1				
≤ 20	37	91	1.0	—
21-40	60	90	1.0	(0.5-1.7)
41-60	90	44	2.5	(1.4-4.7)
> 60	135	31	5.4	(2.8-10.4)
m1/m2				
≤ 20	8	27	0.8	(0.3-1.9)
21-40	19	12	2.1	(0.9-5.0)
41-60	19	4	5.2	(1.5-17.5)
> 60	23	5	5.1	(1.7-15.6)
GSTM1 genotype[e,f]				
Present				
≤ 20	19	56	1.0	—
21-40	31	49	1.1	(0.5-2.3)
41-60	56	23	3.9	(1.7-8.6)
> 60	73	11	9.2	(3.7-22.8)
Null				
≤ 20	27	63	1.3	(0.6-2.7)
21-40	48	53	1.6	(0.8-3.2)
41-60	53	27	3.1	(1.4-6.8)
> 60	88	25	5.2	(2.5-12.2)

[a] Analysis limited to cigarette smokers (395 cases and 307 controls).
[b] Odds ratios (OR) and 95% confidence intervals (CI) adjusted for the other genotype, age, gender, pack-years, and years since quit smoking.
[c] m1/m1 = homozygous wild-type; m1/m2 heterozygous. Homozygous variant subjects (m2/m2) were excluded from this analysis ($n = 8$).
[d] Likelihood ratio test for homogeneity: $\chi^2_{(3)} = 3.8$, $P = 0.3$.
[e] *GSTM1* null genotype refers to homozygous deletion.
[f] Likelihood ratio test for homogeneity: $\chi^2_{(3)} = 4.1$, $P = 0.3$.

cumulative smoking dose and lung cancer risk by the *CYP1A1 MspI* and the *GSTM1* polymorphisms. This analysis was restricted to current and former smokers. The genotype ORs among never-smokers were very similar to the ORs for all subjects reported in Table 3 (OR = 1.8, CI = 0.7-5.0 for *CYP1A1 MspI* heterozygous genotype, and OR = 1.1, CI = 0.4-2.6 for *GSTM1* null genotype). Stratification by *CYP1A1 MspI* suggested a stronger effect of smoking among subjects with the *MspI* heterozygous genotype than among *MspI* wild-type subjects, however, CIs were too wide to enable any inference (likelihood ratio test [LRT] for homogeneity: $\chi^2_{(3)} = 3.8$, $P = 0.3$). Stratification by *GSTM1* genotype suggested a weaker effect of pack-years for subjects with the null genotype than for subjects with *GSTM1* present. However, as in the case of the *CYP1A1* genotype the test for homogeneity indicated that this difference is likely to be explained by chance (LRT for homogeneity: $\chi^2_{(3)} = 4.1$, $P = 0.3$).

In order to improve the power to study these potential interactions and to avoid the dependence of results on arbitrary cut-off points, we also modeled pack-years as a continuous variable. Generalized additive models indicated that the relationship between the logit of the probability of disease and pack-years for the different genotype groups was well-approximated by second-degree polynomials. Therefore, we present results from parametric models which included second-degree polynomials to describe the effect of pack-years by the different genotypes.

Figures 1 and 2 represent the relationship between pack-years and the logit of the probability of disease after stratification by the *CYP1A1 MspI* and *GSTM1* genotypes. Figure 1 shows an about 50 percent increase in risk of lung cancer for subjects with the *MspI* heterozygous genotype, as seen in Table 2, which does not seem to change across levels of cumulative cigarette dose (LRT for interaction: Deviance = -0.02, $P = 1.0$). Thus, the association between cumulative smoking dose and lung cancer does not appear to be modified by the *CYP1A1 MspI* genotype. According to Figure 2, *GSTM1* null subjects seem to have an increased risk of lung cancer at low to moderate doses of smoking. However, as the number of pack-years increases, the lung cancer risk increases at a lower rate among subjects with the *GSTM1* null genotype than among subjects with *GSTM1* present; thus, *GSTM1* null subjects appear to have a decreased risk of lung cancer at high doses of smoking. This pattern of interaction agrees with the pattern observed in the stratified analysis (Table 4), however according to the test for interaction we cannot reject the null hypothesis that the effect of smoking is the same across *GSTM1* genotypes (LRT for interaction: Deviance = -4.1, $P = 0.1$).

Interaction between the GSTM1 and CYP1A1 MspI polymorphisms

When compared with subjects with at least one *GSTM1* allele and with the *CYP1A1 MspI* homozygous wild-type genotype, neither the *CYP1A1 MspI* heterozygous genotype nor the *GSTM1* null genotype was associated with an increase in lung cancer risk; however, having both traits was associated with an approximately twofold increase in risk (CI = 1.0-3.4) (Table 4). Differences between the crude and adjusted estimates shown in Table 4 were due

Figure 1. Fitted values for the relationship between lung cancer risk and pack-years stratified by the *CYP1A1 MspI* polymorphism, from a logistic regression model which included pack-years as a second-degree polynomial and terms for age, gender, years since quitting smoking, and *GSTM1* genotype.

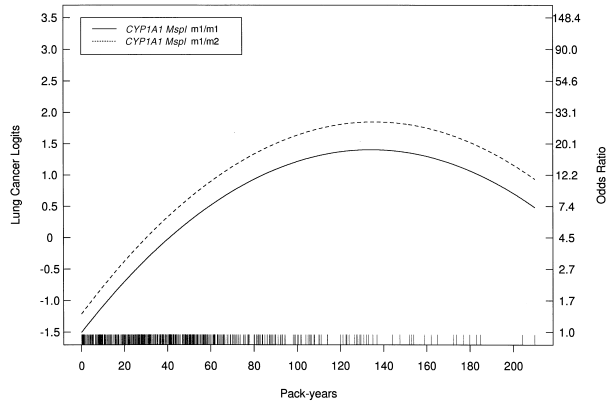

Figure 2. Fitted values for the relationship between lung cancer risk and pack-years stratified by the *GSTM1* polymorphism, from a logistic regression model which included pack-years as a second-degree polynomial and terms for age, gender, years since quitting smoking, and *CYP1A1 MspI* genotype.

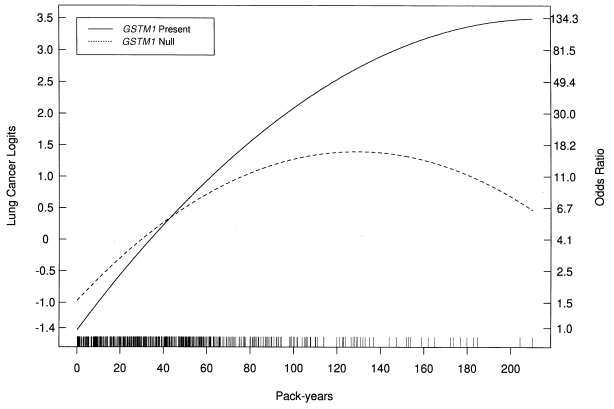

Table 4. Association between *CYP1A1 MspI* and *GSTM1* polymorphisms and lung cancer risk

	Cases	Controls	Crude		Adjusted[b]	
			OR	(CI)[a]	OR	(CI)[a]
CYP1A1 MspI m1/m1[c]						
GSTM1 present	174	155	1.0	—	1.0	—
GSTM1 null	195	182	1.1	(0.8-1.4)	0.9	(0.6-1.3)
CYP1A1 MspI m1/m2[c]						
GSTM1 present	38	32	0.9	(0.6-1.6)	1.0	(0.5-1.9)
GSTM1 null	35	43	1.4	(0.8-2.3)	1.9	(1.0-3.4)
Total	442	412				

[a] OR = odds ratio; (CI) = 95 percent confidence interval.
[b] Odds ratios adjusted for age, gender, pack-years, and years since quit smoking.
[c] m1/m1 = homozygous wild-type; m1/m2 = heterozygous. Homozygous variant subjects (m2/m2) were excluded from this analysis ($n = 8$).
Likelihood ratio test for homogeneity: $\chi^2(1) = 1.7$, $P = 0.1$.

to the inclusion of pack-years in the adjusted model. This reflected the association between *CYP1A1 MspI* genotype and cigarettes smoked per day reported in the previous section.

No important differences in the interaction between the *GSTM1* and *CYP1A1 MspI* genotypes were observed across different histologic types. The adjusted OR for the interaction term between the two genotypes was 2.1 (CI = 0.9-4.0, $P = 0.1$) for all tumors combined, 2.3 (CI = 0.8-6.0, $P = 0.1$) for adenocarcinoma and 2.2 (CI = 0.5-10.0, $P = 0.3$) for squamous cell carcinoma. Thus, although there is an indication of a positive multiplicative interaction between *GSTM1* and *CYP1A1 MspI* polymorphisms, our data do not provide enough evidence to reject the null hypothesis of no multiplicative interaction at the five percent level of significance.

Since only the combination of *GSTM1* null and *CYP1A1 MspI* heterozygous genotype seems to be associated with lung cancer risk, we compared the risk of subjects with this genotype combination with the risk of subjects with any other genotype combination, namely *GSTM1* present/*CYP1A1 MspI* homozygous wild-type, *GSTM1* present/*CYP1A1 MspI* heterozygous or *GSTM1* null/*CYP1A1 MspI* homozygous. The adjusted OR for this comparison was 2.0 (CI = 1.1-3.5) for all tumors combined, 1.8 (CI = 0.9-3.4) for adenocarcinoma, and 1.8 (CI = 0.7-4.5) for squamous cell carcinoma.

Our data suggested a negative interaction between pack-years and the *GSTM1* genotype. Therefore we attempted to explore a potential three-way interaction between the two genotypes and pack-years. Given the limited power of our study to assess a three-way interaction, we constructed a reduced model to explore the modification of the OR for the combined genotype by two levels of cumulative smoking dose using 40 pack-years as the cut-off point (data not shown). This reduced model provided no evidence of a substantial modification of the pack-years effect by the combined genotype (LRT for homogeneity: $\chi^2(1) = 0.1$, $P = 0.8$).

Discussion

In this North American, mostly White, population we did not find evidence for an overall association between the *GSTM1* null genotype and lung cancer risk. However, our data suggested that the *GSTM1* deletion may contribute to lung cancer risk in combination with the *CYP1A1 MspI* heterozygous variant genotype. The frequency of the *GSTM1* null genotype among White controls (227/436 = 52 percent) was very similar to the only previously published study in North American Whites (244/465 = 52 percent)[20] and it was within the range of frequencies found in other White populations: 53 percent (175/329) in a Swedish population,[25] 53 percent (82/155) in a German population,[35] 44 percent (62/142) in a Finnish population,[23] 42 percent (94/225) in a English population,[17] and 46 percent (67/147) in a northwestern Mediterranean population.[21] These variations might reflect differences in ethnicity within Caucasian populations.[36] However, they also could reflect random variations due to the small samples in many of these studies.

As in our previous report,[27] the *CYP1A1 MspI* heterozygous genotype was associated with an increased risk of lung cancer, which was evident only after adjusting for pack-years (OR = 1.5, CI = 1.0-2.3). This association did not appear to be modified by the level of cigarette smoking. The crude OR for the *GSTM1* null genotype (OR = 1.1, CI = 0.8-1.4) was similar to the estimate obtained in a recent metanalysis of past studies in Whites (OR = 1.2, CI = 1.0-1.4)[19] and to the crude estimate observed by London *et al*[20] among North American

Whites (OR = 1.0, CI = 0.7-1.3, calculated from Table 2 of the paper). We did not find important differences between adenocarcinoma and squamous cell carcinoma. This finding agrees with a recent meta-analysis[19] and the London *et al* study.[20] Overall, the London *et al* study and the present study indicate that the *GSTM1* null genotype is not associated with a substantive increase in lung cancer risk in North American White populations.

Several epidemiologic studies which have evaluated the interaction between *GSTM1* and cumulative cigarette smoking have obtained contradictory results. Out of the six studies which evaluated this interaction, three[22-24] found a positive association between *GSTM1* and lung cancer only among smokers with high cumulative cigarette dose, two[11-20] found this association only among smokers of low cumulative dose, and one study[35] found no evidence for differences in two categories of cumulative smoking dose. A stronger association between *GSTM1* null genotype and lung cancer for heavier than lighter smokers is also supported by a study[37] which found an association between high sister chromatic exchange (SCE) levels and the *GSTM1* deficiency among subjects with high cotinine levels but not among subjects with lower cotinine levels. Our data suggested a weaker effect of pack-years for *GSTM1* null subjects than for subjects with the *GSTM1* present, which implied a positive association between *GSTM1* null genotype and lung cancer for low to moderate doses of cigarette smoke, and a negative association for high doses of smoking. However, given the biological implausibility of a cross-over effect, chance seems a reasonable explanation for the suggested pattern of interaction.

When we stratified subjects according to both the *CYP1A1 MspI* and *GSTM1* genotypes, while neither the *GSTM1* null genotype alone nor the *CYP1A1 MspI* heterozygous genotype alone were associated with a significant increase in lung cancer risk, having both genetic traits was associated with an approximately twofold increase in risk (CI = 1.0-3.4). This multiplicative interaction was similar for adenocarcinoma and squamous cell carcinoma cases. A similar interaction was reported in a Swedish study of all tumor types[25] but not in two Japanese studies of squamous cell and small cell carcinomas[11,38] which found a positive multiplicative interaction with the *CYP1A1 MspI* homozygous variant genotype but not with the heterozygous genotype.

Several potential biases might have influenced our results. First, our control population includes friends or spouses referred by cases, by cardiac patients, or by thoracic patients. A necessary condition for the validity of our estimates of the genotype association is that, conditional on the variables controlled in the analysis, (i) the genotype distribution of the three controls series is the same, and (ii) represents the distribution in the source population. To evaluate whether (i) is satisfied, we tested for differences in the genotype frequencies of the three control groups conditional on the variables included in our final model, finding no significant differences ($\chi^2_{(2)}$ = 4.4, P = 0.1 for *GSTM1* genotype; and $\chi^2_{(2)}$ = 4.9, P = 0.1 for *CYP1A1 MspI* genotype). Given that we found no significant differences in genotype frequencies and that the frequencies in the collapsed control group were similar to the estimates observed in past studies of White populations, it is seems acceptable to assume that the three control groups represent a common population.

This still leaves us with the possibility that the three control groups, while originating from the same population, might not represent the source population of the cases. Selecting controls among friends or spouses of lung cancer patients or other surgical patients might introduce bias by two different mechanisms: (i) the distribution of exposures among 'gregarious subjects,'[39] *i.e.*, subjects who tend to be named by more than one other person as controls, might be different than among 'non-gregarious subjects' and, therefore, it might not represent the distribution in the source population which includes both types of subjects;[39] (ii) exposures among friends and spouses of patients might be similar to the exposures among the patients themselves, and thus it might not represent the exposure distribution in the source population. These two factors are likely to be relevant for exposures such as age, gender, smoking, or dietary habits. However, within categories of these variables and among the same racial or ethnic group, metabolic polymorphisms are unlikely to be associated with being a gregarious individual or with having a friend or spouse with lung cancer. Therefore, the genotype-disease association is unlikely to be affected substantially by this type of bias. On the other hand, the reported estimates for the effect of pack-years on lung cancer are likely to be underestimated.

Second, our case-population includes only surgical lung cancer patients (stages I and II), and more advanced tumors not eligible for surgery are not included. Thus, differential selection of cases with respect to the genotype could occur if the genotype is related to the stage at which the lung cancer is detected. Although to our knowledge this has not been evaluated for *GSTM1* in past studies, two Japanese studies suggested a higher frequency of the *CYP1A1 MspI* variant allele for poorly differentiated than for well-differentiated adenocarcinomas[40] and for lung cancer patients with metastasis than for patients without metastasis.[41] These differences could have introduced underestimation of the genotype ORs because poorly differentiated tumors and tumors with metastasis are more likely to become inoperable.

Third, ethnicity, which was not measured in this study, could be a confounder of the ORs for the genotype-disease association. However, in the context of our study,

this type of confounding is unlikely to be important since in this almost exclusively White population, ethnicity is unlikely to be strongly associated either with the genotype, as indicated above, or with lung cancer risk.

On the basis of this discussion, we conclude that among Whites, the combination of the *CYP1A1 MspI* heterozygous genotype and the *GSTM1* null genotype is likely to be associated with an increased risk of lung cancer. This indicates that host susceptibility to lung cancer may depend on the metabolic balance between the *P450 1A1* and the *GSTM1* enzyme activities. Our data did not provide with enough evidence for a modification of the effect of pack-years on lung cancer risk by the *CYP1A1 MspI* and *GSTM1* genotypes, however this result is tentative in view of the limited statistical power to assess comprehensibly a potential three-way interaction between *CYP1A1 MspI*, *GSTM1*, and lung cancer risk.

Acknowledgements — The authors thank the following members of the lung cancer susceptibility project: Ms Linda Lineback for patient recruitment; Ms Lucy Ann Principe-Hasan and Mr Nick Weidemann for data entry and management; Ms Marcia Chertok for project coordination; Ms Marlys Rogers and Ms Lucille Pothier for computer programming; Ms Zheng-fa Zuo for PCR analysis; Ms Debby Strahs for sample handling and preparation; and Drs H. Grillo, Douglas Mathiesen, and H. Kazemi of Massachusetts General Hospital. We also thank Drs David Hunter, Richard Monson, James Robins, and Dimitrios Trichopoulos for their critical review of the manuscript.

References

1. Amdur MO, Doull J, Klaassen CD. *Toxicology*. New York, NY (USA): McGraw-Hill, Inc., 1991.
2. Ketterer B, Harris JM, Talaska G, *et al.* The human glutathione S-transferase supergene family, its polymorphism, and its effect on susceptibility to lung cancer. *Environ Health Persp* 1992; 98: 87-94.
3. Kadlubar FF, Hammons GJ. Role of cytochrome P-450 in metabolism of chemical carcinogens. In: Guengerich FP, ed. *Mammalian Cytochromes P-450. Volume 2*. Boca Raton FL (USA): CRC Press, 1987: 81-130.
4. Peterson DD, McKinney CE, Ikeya K, *et al.* Human CYP1A1 gene: Cosegregation of the enzyme inducibility phenotype and a RFLP. *Am J Hum Genet* 1991; 48: 720-5.
5. Hayashi S, Watanabe J, Nakachi K, Kawajiri K. Genetic linkage of lung cancer associated MspI polymorphisms with amino acid replacement in the heme binding region of the human cytochrome P450 1A1 gene. *J Biochem* 1991; 110: 407-11.
6. Cosma G, Crofts F, Taioli E, Toniolo P, Garte S. Relationship between genotype and function of the human CYP1A1 gene. *J Toxicol Environ Health* 1993; 40: 309-19.
7. Crofts F, Taioli E, Trachman J, *et al.* Functional significance of different human CYP1A1 genotypes. *Carcinogenesis* 1994; 15: 2961-3.
8. Nakachi K, Imai K, Hayashi S, Watanabe J, Kawajiri K. Genetic susceptibility to squamous cell carcinoma of the lung in relation to cigarette smoking dose. *Cancer Res* 1991; 51: 5177-80.
9. Kawajiri N, Nakachi K, Imai K, Yoshii A, Shinoda N, Watanabe J. Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P4501A1 gene. *FEBS* 1990; 263: 131-3.
10. Hayashi S, Watanabe J, Kawajiri K. High susceptibility to lung cancer analyzed in terms of combined genotypes of P450IA1 and Mu-class glutathione S-transferase genes. *Jpn J Cancer Res* 1992; 83: 866-70.
11. Nakachi K, Imai K, Hayashi S, Kawajiri K. Polymorphisms of the CYP1A1 and the glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. *Cancer Res* 1993; 53: 2994-9.
12. Drakoulis N, Cascorbi I, Brokmoller J, Gross CR, Roots I. Polymorphisms in the human CYP1A1 gene as susceptibility factors for lung cancer: exon-7 mutation (4889 A to G), and a T to C mutation in the 3'-flanking region. *Clin Investig* 1994; 72: 240-8.
13. Hirvonen A, Husgafvel-Pursiainen K, Karjalainen A, Antila S, Vainio H. Point-mutational MspI and Ile-Val polymorphisms closely linked in the CYP1A1 gene: lack of association with susceptibility to lung cancer in a Finish study population. *Cancer Epidemiol Biomark Prev* 1992; 1: 485-9.
14. Shields PG, Caporaso NE, Falk RT, *et al.* Lung cancer, race, and a CYP1A1 genetic polymorphism. *Cancer Epidemiol Biomark Prev* 1993; 2: 481-5.
15. Sugimura H, Hamada GS, Suzuki I, *et al.* CYP1A1 and CYP2E1 polymorphism and lung cancer, case-control study in Rio de Janeiro, Brazil. *Pharmacogenetics* 1995; 5: S145-8.
16. Seidegård J, Vorachek WR, Pero RW, Pearson WR. Hereditary differences in the expression of the human glutathione transferase active on trans-stilbene oxide are due to a gene deletion. *Proc Natl Acad Sci* 1988; 85: 7593-7.
17. Zhong S, Howie AF, Ketterer B, *et al.* Glutathione S-tranferase mu locus: use of genotyping and phenotyping assays to assess association with lung cancer susceptibility. *Carcinogenesis* 1991; 12: 1533-7.
18. Brockmöller J, Gross D, Kerb R, Dakoulis N, Roots I. Correlation between *trans*-stilbene oxide-glutathione activity and the deletion mutation in the glutathione S-transferase class Mu gene detected by polymerase chain reaction. *Biochem Pharmacol* 1992; 43: 650-4.
19. McWilliams JE, Sanderson BJS, Harris EL, Richert-Boe KE, Henner WD. Glutathione S-transferase M1 (GSTM1) deficiency and lung cancer risk. *Cancer Epidemiol Biomark Prev* 1995; 5: 589-94.
20. London SJ, Daly AK, Cooper J, Navidi WC, Carpenter CL, Idle JR. Polymorphism of glutathione S-transferase M1 and lung cancer risk among African-Americans and Caucasians in Los Angeles County, California. *JNCI* 1995; 87: 1246-53.
21. To-Figueras J, Gené M, Gómez-Catalán J, *et al.* Gluthathione S-transferase M1 and codon 72 P53 polymorphism in a Northwestern Mediterranean population and their relation to lung cancer susceptibility. *Cancer Epidemiol Biomark Prev* 1996; 5: 337-43.
22. Nazar-Stewart V, Mutulsky AG, Eaton DL, *et al.* The glutathione S-transferase mu polymorphisms as a marker

for susceptibility to lung cancer. *Cancer Res* 1993; 53(Suppl): 2313-8.
23. Hirvonen A, Husgafvel-Pursiainen K, Anttila S, Vainio H. The GSTM1 null genotype as a potential risk modifier for squamous cell carcinoma of the lung. *Carcinogenesis* 1993; 14: 1479-81.
24. Kihara M, Khiara M, Noda K. Lung cancer risk of GSTM1 null genotype is dependent on the extent of tobacco smoke exposure. *Carcinogenesis* 1994; 15: 415-8.
25. Alexandrie AK, Sundberg MI, Seidegård J, Tornling G, Rannug A. Genetic susceptibility to lung cancer with special emphasis on CYP1A1 and GSTM1: a study on host factors in relation to age at onset, gender and histological cancer types. *Carcinogenesis* 1994; 15: 1785-90.
26. Anttila S, Hirvonen A, Husgafvel-Pursiainen K, Karjalainen A, Nurminen T, Vainio H. Combined effect of CYP1A1 inducibility and GSTM1 polymorphism on histological type of lung cancer. *Carcinogenesis* 1994; 15: 1133-5.
27. Xu X, Kelsey K, Wiencke J, Wain J, Christiani D. Cytochrome P450 CYP1A1 MspI polymorphism and lung cancer susceptibility. *Cancer Epidemiol Biomark Prev* 1996; 5: 687-92.
28. Ferris BG. Epidemiology standardization project. *Am Rev Respir Dis* 1978; 118(Part 2): 1-120.
29. Willett WC, Sampson L, Stampfer MJ, *et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985; 122: 51-65.
30. García-Closas M, Kelsey KT, Wiencke JK, Wain JC, Christiani DC. Dietary intake as a modifier of the association between GSTM1 and CYP1A1 MspI polymorphisms and lung cancer susceptibility. Doctoral Thesis, Chapter 3. Harvard School of Public Health, Boston, MA, 1996.
31. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 1991; 10: 506-13.
32. Chen H, Sandler DP, Taylor JA, *et al.* Increased risk for myelodysplastic syndromes in individuals with glutathione transferase theta 1 (*GSTT1*) gene defect. *Lancet* 1996; 347: 295-6.
33. Cheng TJ, Christiani DC, Xu X, Wain JC, Wiencke JK, Kelsey KT. Glutathione S-transferase μ genotype, diet, and smoking as determinants of sister chromatid exchange frequency in lymphocytes. *Cancer Epidemiol Biomarkers Prev* 1995; 4: 535-42.
34. Hastie TJ, Tibshirani RJ. *Generalized Additive Models.* London, UK: Chapman & Hall, 1990:
35. Brockmöller J, Kerb R, Drakoulis N, Nitz M, Roots I. Genotype and phenotype of glutathione S-transferase class μ isoenzymes and in lung cancer patients and controls. *Cancer Res* 1993; 53: 1004-11.
36. Lin HJ, Han C-Y, Bernstein DA, Hsiao W, Lin BK, Hardy S. Ethnic distribution of the glutathione transferase Mu 1-1 (GSTM1) null genotype in 1473 individuals and application to bladder cancer susceptibility. *Carcinogenesis* 1994; 15: 1077-81.
37. Poppel GV, Vogel NCL Balderen PJV, Kok FJ. Increase cytogenetic damage in smokers deficient in glutathione S-transferase isoenzyme mu. *Carcinogenesis* 1992; 13: 303-5.
38. Kihara M, Kihara M, Noda K. Risk of smoking for squamous and small cell carcinomas of the lung modulated by combinations of CYP1A1 and GSTM1 gene polymorphisms in a Japanese population. *Carcinogenesis* 1995; 16: 2331-6.
39. Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. *Epidemiology* 1990; 1: 273-84.
40. Nakachi K, Hayashi S-i, Kawajiri K, Imai K. Association of cigarette smoking and CYP1A1 polymorphisms with adenocarcinoma of the lung by grades of differentiation. *Carcinogenesis* 1995; 16: 2209-13.
41. Okada T, Kawashima K, Fukushi S, Minakuchi T, Nishimura S. Association between a cytochrome P450 CYP1A1 genotype and incidence of lung cancer. *Pharmacogenetics* 1994; 4: 333-40.